OPTIMAL LOCATION FOR A HEALTH FOOD RESTAURANT

Bhavvya Sarna 13th April 2021

Introduction

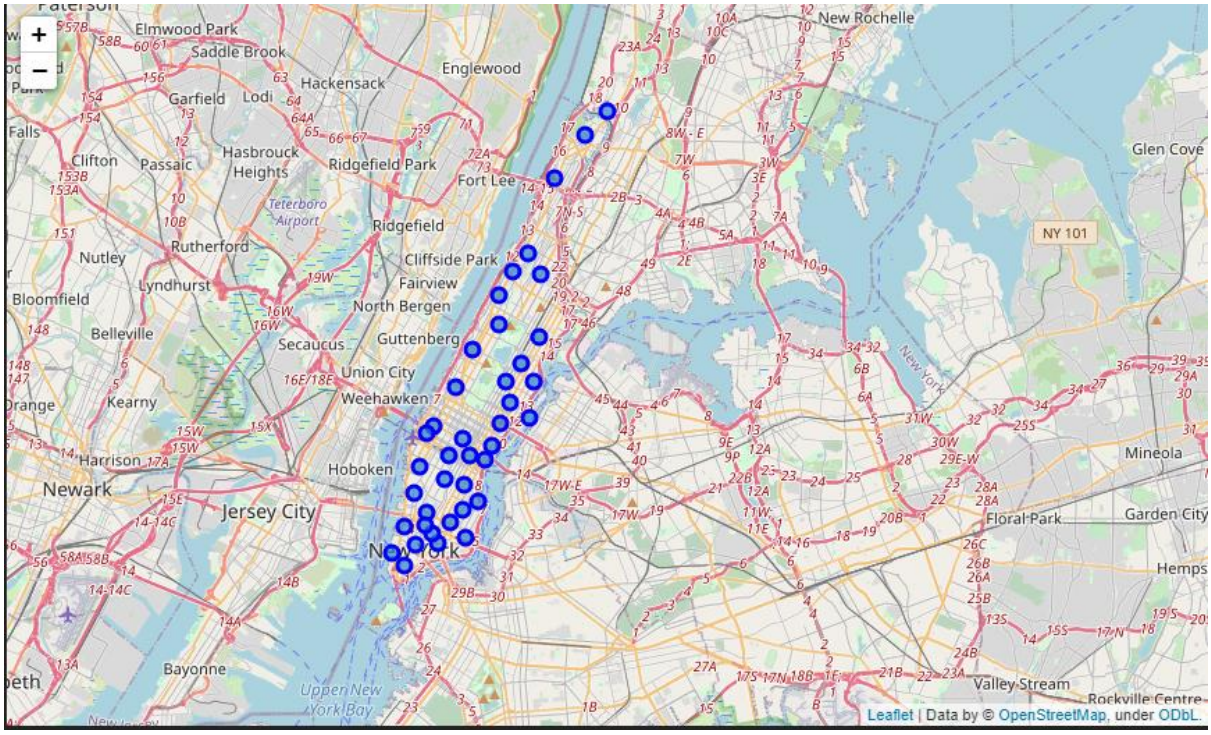
A large scale social shift towards a sedentary lifestyle has led to a peak in obesity rates, increased risk of cardiovascular diseases, and rampancy of associated mental health issues such as depression. An increasing number of people have begun to recognize the looming threat of deteriorating fitness levels, especially in developed regions. Many have adopted healthier eating habits in order to combat this issue. An increase in demand for healthy eating options has led to the opening of numerous health food restaurants. This project aims at determining the best neighborhood in Manhattan for a restaurant owner to open a healthy eating focused restaurant.

Data

This project makes use of New York borough and neighborhood location data provided by IBM for the Data Science Course in json format. From “features” in the json file, names of boroughs, neighborhoods in New York, along with their corresponding latitude and longitude were extracted. Data about venues in each neighborhood has been obtained via the foursquare API using the latitude and longitude of each neighborhood.

Methodology

First, all the neighborhoods and their location data were collected in a dataframe. Since this project focuses on Manhattan specifically, only the neighbourhoods in Manhattan were considered from here on as identified in the map below.



For each neighborhood, upto 30 venues within a 500 meter range were obtained using the foursquare API. The same was used to find the category to which each venue belonged. For example, the first venue is “Arturo’s”, a pizza place in Marble Hill.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Marble Hill	40.876551	-73.91066	Arturo's	40.874412	-73.910271	Pizza Place
1	Marble Hill	40.876551	-73.91066	Bikram Yoga	40.876844	-73.906204	Yoga Studio
2	Marble Hill	40.876551	-73.91066	Tibbett Diner	40.880404	-73.908937	Diner
3	Marble Hill	40.876551	-73.91066	Astral Fitness & Wellness Center	40.876705	-73.906372	Gym
4	Marble Hill	40.876551	-73.91066	Dunkin'	40.877136	-73.906666	Donut Shop

The target demographic of the client is people who are conscious about their health and lifestyle since they would be the ones most likely to become regular customers. However, too many restaurants in an area would lead to increased competition and loss of potential customers. So relevant venue categories were selected and grouped under umbrella terms like “Restaurants” which includes restaurants of various cuisines and “gyms” which includes all types of gyms and fitness centers. Then, the total number of such categories was counted for each neighborhood, resulting in the table of which a preview is shown below:

	Neighborhood	gyms	restaurants	Health Food Store	Sporting Goods Shop	Farmers Market
0	Battery Park City	1	1	0	0	0
1	Carnegie Hill	3	6	0	0	0
2	Central Harlem	1	8	0	0	0
3	Chelsea	0	10	0	0	0
4	Chinatown	0	9	0	0	0

These values were normalized using the maximum value in each column. The normalized values were used for a K-Means Clustering algorithm, which categorized each neighborhood into one of 5 clusters based on their similarity to one another.

Result

The mean of normalized value counts of various venue categories in each cluster can be used to make recommendations to the client.

labels	gyms	restaurants	Health Food Store	Sporting Goods Shop	Farmers Market
0	0.063158	0.673684	0.0	0.000000	0.0
1	0.550000	0.433333	0.0	0.000000	0.0
2	0.100000	0.233333	0.0	0.833333	0.0
3	0.160000	0.386667	0.0	0.000000	1.0
4	0.900000	0.300000	1.0	0.166667	0.0

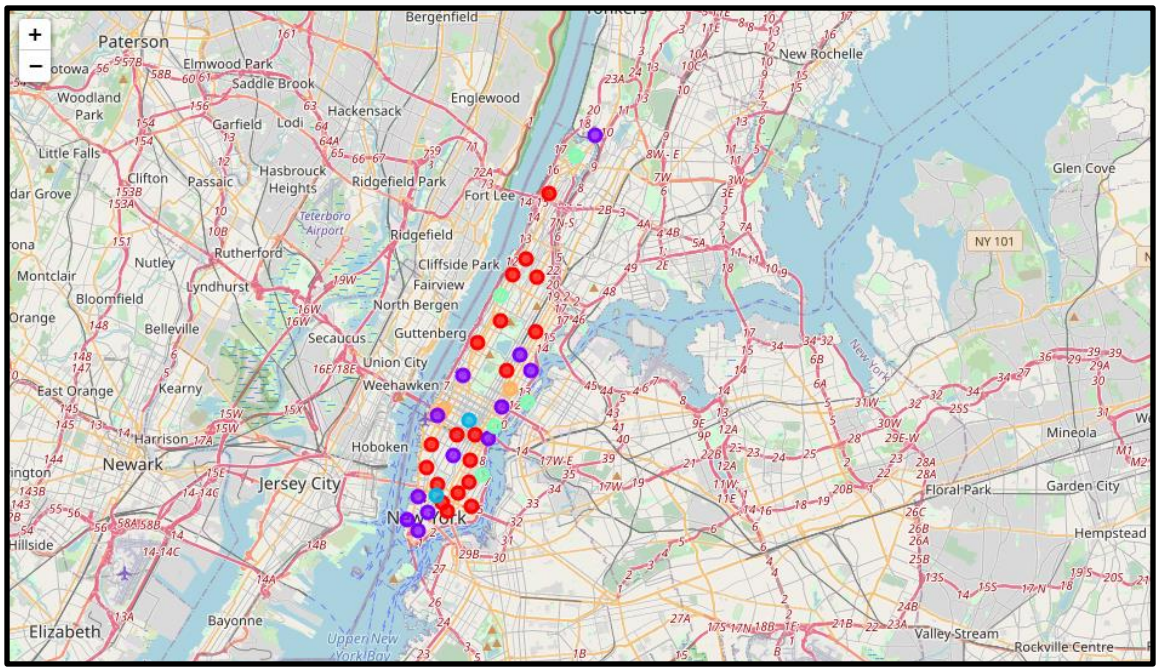
However, this data isn't particularly intuitive. So these values can be multiplied with the original maximum of each column to get an expected number of venues belonging to each category for each cluster.

labels	gyms	restaurants	Health Food Store	Sporting Goods Shop	Farmers Market
0	0.315789	10.105263	0.0	0.0	0.0
1	2.750000	6.500000	0.0	0.0	0.0
2	0.500000	3.500000	0.0	2.5	0.0
3	0.800000	5.800000	0.0	0.0	1.0
4	4.500000	4.500000	1.0	0.5	0.0

Based on the expected values, neighborhoods with the label 4 would be the best fit for the client since they have the maximum number of gyms and not too many other restaurants. The presence of Sporting good shops and health food stores further indicates that these neighborhoods would be home to people fitting the target demographic.

An alternative to this could be a neighborhood in cluster 2 since it clearly has a very sport focused environment. Coupled with the fact that cluster 2 neighborhoods also have the least number of competing restaurants, they also make suitable candidates for the client to open their restaurant.

The neighborhoods after clustering look as follows:



Legend

Custer 0 - Red

Cluster 1 - Purple

Cluster 2 - Blue

Cluster 3 - Green

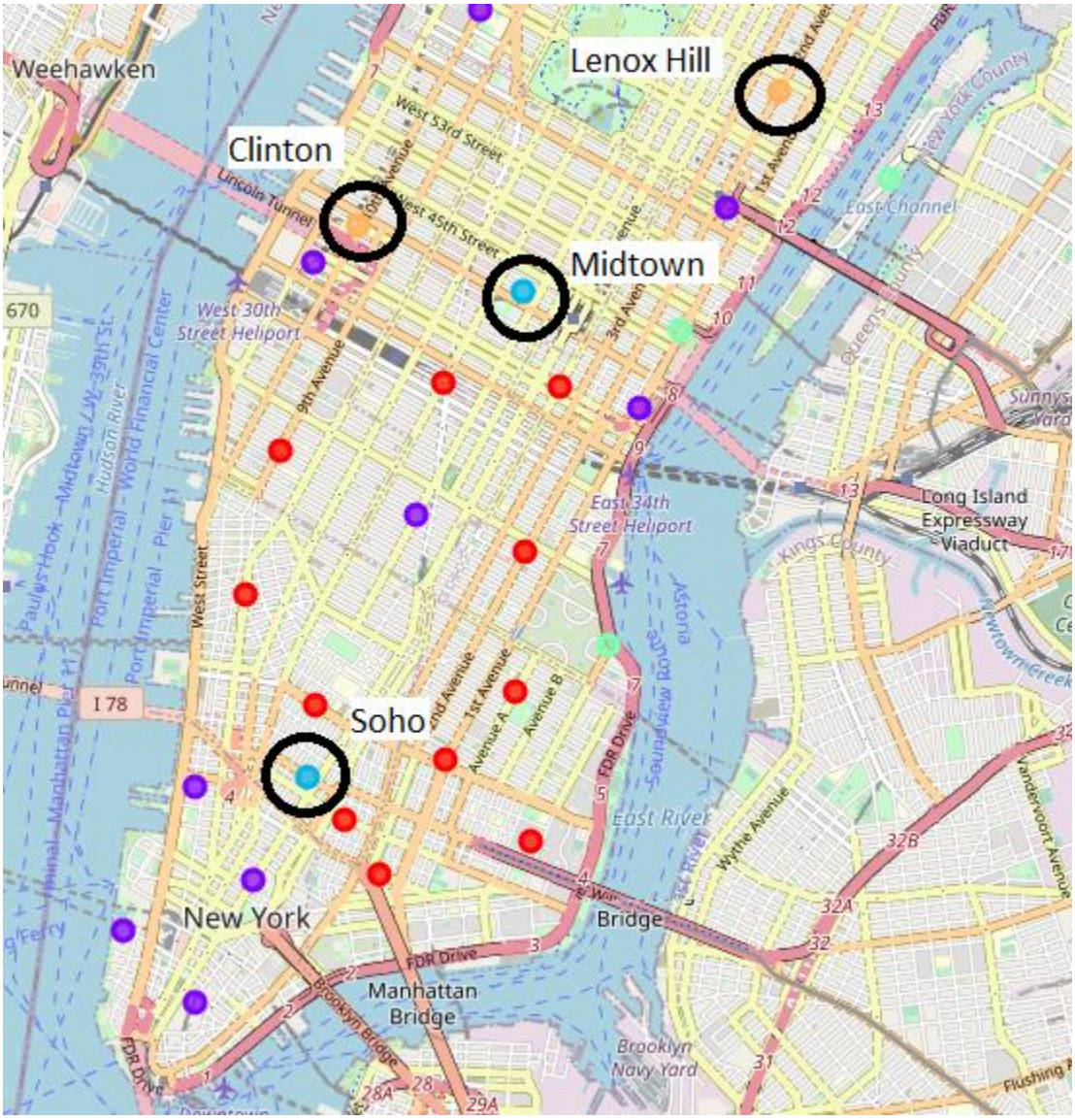
Cluster 4 - Orange

Discussion

On the basis of the clustering analysis, the neighbourhoods belonging to clusters 2 and 4 would be recommended.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	label
6	Clinton	40.759101	-73.996119	4
16	Lenox Hill	40.768113	-73.958860	4
23	Midtown	40.754691	-73.981669	2
29	Soho	40.722184	-74.000657	2

Thus, the client is advised to open their restaurant in Clinton or Lenox Hill. If neither of these is feasible, then the secondary recommendations would be Midtown or Soho.



Conclusion

This project has analyzed data about neighbourhoods in Manhattan and identified 4 candidate neighborhoods wherein a health food restaurant can be opened to maximize profits based on demographics. The final decision, of course, would also depend on factors such as location-based expenses and supply chains logistics.